Quantitative Qualitative Correspondence in Grammaticalization

Abigail Petrsoric (abigail.petrsoric@uconn.edu)

Hyosun Lee (hyosun.lee@uconn.edu)

Whitney Tabor (whitney.tabor@uconn.edu)

Department of Psychological Sciences-U1020
University of Connecticut
Storrs, CT 06269-1020 USA

Abstract

The gradual nature of historical language change is widely acknowledged. We explore a syntactic change model that offers a new view into the theoretical difference between classical and neural network claims about language encoding. Most prior treatments of *grammaticalization* fail to account for how, exactly, new forms arise, focusing instead on change following innovation. A phenomenon relevant to the innovation puzzle is Quantitative Anticipation of Qualitative Change in Grammaticalization (QAQCG)—gradual statistical changes anticipate structural changes. Although prior researchers have given phenomenological descriptions, we know of no rigorous method for testing whether QAQCG exists. Here, we quantitatively examine the case of English *a lot* which has grammaticalized an Adverb function from a Noun Phrase function. A simple feedforward neural network implements QAQCG, predicting a curving trajectory in probability space. Bayes Factor analysis supports the network over a classically-motivated linear model, highlighting continuity and nonlinearity as distinctive theoretical claims.

Introduction

The morpho-syntactic structure of a human language tends to evolve, in some cases quite radically, over periods of 100s to 1000s of years. We focus on “grammaticalization" (Meillet, 1912)—content morphemes like nouns, verbs, and adjectives transform into function morphemes—prepositions, complementizers, adverbs, etc. [1] Examples of such changes include the development of verbs of saying into complementizers (e.g., *that*) (Lord, 1976), the development of nominal emphasizers (e.g., *a crumb*, *a step*) into negation markers (Jespersen, 1917), the development of motion verbs (e.g., *go*) into future markers (J. Bybee et al., 1991).

We use the word “evolves" to emphasize the fact that, intuitively, such changes happen gradually. The gradualness is apparent in several ways. For one, native speakers do not generally experience any radical break in communicative function; indeed, many grammatical shifts are not even noticed by speakers unless attention is explicitly drawn to them. Also, there is often a semantic change associated with the syntactic change and the semantic change is generally very subtle. For example, the established comitative verb/preposition element, *bii*, in the Austronesian language To'aba'ita (1a) has an emerging conjunction usage (1b) (Lichtenberk, 1991). The semantic shift from WITH to AND is small in the sense that the distinction between someone participating in an activity (WITH-meaning) and contributing to the activity (AND-meaning) is a very subtle distinction to draw.

(1) a.

Nau	*kwai*	*lai*	**bii**	*kera*
I	I :IMPFV	go	WITH	them

‘I will go with them.’

b.

Tha	*Maeli*	**bii**	*nau*	*mera*	*lae*
ART	Maeli	AND	I	we(DU,EXCL) :PFV	go
'i	*ro'o*	*ura*	*dee-laa*		
in	yesterday	for	fish-NOM		

‘Maeli and I went fishing yesterday.’

Additionally, in some cases, there is an accompanying phonological change which is also generally gradual, often involving attrition or mutation (e.g, PGmc *mati* (‘food’) + *sahsa* (‘cutting device’) > Dutch *mes* (‘knife’), Van Bree, 2004; Classical Latin *cantare habeo* (‘I have to sing’) > French *chanterai* (‘I will sing’), Lausberg, 1972).

The morpho-syntactic part of the picture is more puzzling : current, explicitly articulated models of the mental encoding of morpho-syntax generally employ symbol combination—indeed, recursive discrete symbol combination seems to be a key ingredient of the expressive power of natural languages. On the face of it, symbol combinations do not seem suited to undergoing gradual metamorphosis.

Nevertheless, there are signs that, in some sense, syntactic forms also change gradually. Minimalist and Cartographic theories of syntax posit a functional hierarchy with a succession of recursively embedded levels (Chomsky, 1995; Cinque, 1999). Roberts & Roussou (2003) provide evidence that many grammaticalization episodes involve progression of morphemes stepwise up the functional hierarchy. Pur-

1. There seems to be a natural class of such changes that is broader than just instances of content → function. It includes changes of one functional type to another (Hopper & Traugott, 1993; Roberts, 2010; Roberts & Roussou, 2003; Traugott, 2017) rare cases that go in the opposite direction structurally (Norde, 2009; Diertani, 2011), and cases of transition from content → content that gain functionality in the sense of abstraction (Sweetser, 1990; J. L. Bybee et al., 1994). In the present work, we are interested in this whole class of changes.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

suing this theme, Roberts (2010) notes that the Cartographic functional hierarchy has a great many levels, with very fine-grained distinctions between them, so if grammaticalized elements were to climb incrementally up the hierarchy, one could argue that their syntax was changing almost continuously.

We find at least some of these observations compelling—the movement upward is uncontroversial over a large natural subclass of grammaticalization examples. However, there is an aspect of grammaticalization's gradualness which this hierarchy-climbing treatment does not address : it has a statistical dimension.

Statistical Gradualness

First, the arrivals of new constructions in corpora have a sigmoidal form : low frequencies gradually increase over a period of time, followed by a sharp ramping-up, then a leveling off. Moreover, if we identify the time of first clear evidence in a corpus of a novel grammatical usage, we can often observe statistical changes in already licensed forms, from decades to centuries preceding the moment of advent, that intuitively "prepare the ground" for the appearance of the new form. For example, Andersen (1987) reports on person-number markers in Polish, which have behaved as (2nd-position or "Wackernagel") clitics for several centuries (2a). In some dialects of modern Polish, however, when the descendants of these person-number markers attach to a verb (2b), they trigger a stress-change rule that is strictly associated with word-internal phonological processes (*przsys 'edł-em* ('I arrived') vs. *prz 'zysed-ł* ('arrived')). This suggests that the modern forms are no longer clitics but have become affixes (across languages, affixes function as parts of the words they are attached to, unlike clitics, which are more independent, and can generally attach to a variety of word types). Tracking the frequencies of placement of these person-number markers over four centuries preceding the appearance of evidence for lexical status, Andersen finds that the affixes-to-be appeared increasingly often adjacent to the verb. In other words, the progressive statistical development changed the distribution so that the clitics behaved increasingly like affixes before they actually became affixes. We refer to this kind of phenomenon as *Quantitative Anticipation of Qualitative Change in Grammaticalization* (QAQCG). (see also Craig, 1991; De Smet, 2012; Tabor, 1995; Schmid, 2020).

(2) a. *To*-**m** *jest* *ogla* *dała*
That-1sg EMPH look give

'That I did see'

b. *Wczoraj* *przyszedł*-**em**
Yesterday arrived-1sg

'I arrived yesterday.'

An essential claim of this view is that grammaticalization happens micro-construction by micro-construction. That is, as a series of successive S-curves. Moreover, it is arguable that the sequence of constructions involved is governed by a Proximity Principle : each successive construction in the sequence is maximally similar to the preceding and following types. One commonly occurring type of proximity involves ambiguity : in a historical sequence of the form, A $\rightarrow$ B $\rightarrow$ C, B is often ambiguously analyzable as the A type or the C type, whereas A and C do not overlap in this fashion. For example in tracing the recent development of adjectival use of English *fun*, which was strictly a noun before the 19th century, De Smet (2012) provides partial evidence that the S-curve for rate of occurrence per million words rose first for unambiguously nominal constructions (3a), then for ambiguous Noun-Adjective constructions (3b) and then for unambiguously adjectival constructions (3c)—we have confirmed this characterization by additional tabulation of data from the Corpus of Historical American English (COHA) (Davies, 2008), which De Smet also drew on. De Smet refers to this Proximity Principle as "sneakiness"—the language always makes the least obtrusive changes first, sidling into its radical grammatical divergences.

(3) a. 1751 "Don't mind me tho'–For all my *fun* and jokes." (OED) . [Nominal—Unambiguous]

b. 1898 Wouldn't that be *fun*, Bess. (COHA) [Predicate Nominal or Adjective—Ambiguous]

c. 1995 The tiny downtown is still a *fun* place to stroll (COHA) [Attributive Adjective—Unambiguous]

Observations of QAQCG and sneakiness suggest that there is a meaningful relationship between the frequency changes and the grammatical developments. However, most of the work on the topic does not, in any rigorous way, assess the significance of observed statistical trends or formally evaluate the hypothesis that the statistical changes are meaningfully related to the grammatical developments. Our purpose in the current paper is to establish a plausible and replicable method of making such a formal assessment. We proceed as follows. First we give a recipe for what to count in a corpus. Then, we identify two opposing theses about how the data from the counting should behave. The first thesis, called *Linear Interpolation* functions as a kind of baseline hypothesis. The second thesis, *Higher Order Interpolation* employs a very simple neural network that predicts QAQCG and proximity effects. Then we do model comparison to decide which thesis fits the data better. In the remainder of this paper, we will describe this procedure and then apply it to a case study of the grammatical development of English *a lot* from the early 19th century to the 21st century, during which time it developed from a noun phrase (determiner + noun) into an adverb similar to *much*.

What to Count

Most corpus-based historical treatments study the language in successive batches, each drawn from a relatively short time period (e.g., a decade) and count rates of occurrence relative to the size of the batch. This is a way of asking how prevalent a word is in the whole language. But words and phrases have very restricted distributional affordances. For example, English noun phrases primarily occur as subjects, objects, and prepositional objects ; adverbs typically occur next to adjectives in adjective phrases, and adjacent to verb phrases and clauses. When neural networks learn encodings of words and phrases, and show evidence of mastering the grammatical constraints of a language (Elman, 1990; Devlin et al., 2014; Roumeliotis & Tselikas, 2023), they are tracking the rates of occurrence of adjacent sequences of words and using the distribution over these event-types to profile each expression of interest. Relatedly, we propose using linguistic insight to define structural environments (e.g., for *a lot*, we use the tagged COHA corpus to identify Noun Phrase uses, Quantifier uses, and Adverb uses) and we track the relative frequency of the target expression across these environments. Thus a linguistic collocation is profiled as a probability vector.

To be thorough, for a given language, the profile vector might have many dimensions (as many as there are grammatically distinct environments in the language). However, for a given change, many of these contexts are not of interest because the target expression never interacts with them. Also, it is possible to lump contexts together into more general contexts—for example, *a lot* can function as a noun phrase referring to a large quantity with or without a following *of*-phrase, yet with the same collection or substance understood in both cases (e.g., *ate a lot of food* vs. *ate a lot*). In such a case, the two environment can be grouped together into a super environment. The theory holds that, as long as the groupings correspond to natural grammatical classes and the old and new class distributions are not co-linear, then Higher Order Interpolation should fit the data better than Linear Interpolation.

(4) gives examples of the three major environmental types that we considered for *a lot*.[2]

(4) a. 1850 there is *a lot* of apothecary's stuff aboard, which I traded. . .. (COHA) [Noun Phrase]

b. 1912 We will learn *a lot* of duets together when I come. (COHA) [Quantifier]

c. 1961 Your cruiser's a lot slower than the Mooncat (COHA) [Adverb]

If we are studying a case with just three contexts of interest, we can make a two-dimensional portrayal of the whole probability simplex. Under this portrayal, statistical change in the behavior of the focus expression is modeled by a trajectory in the simplex (Figure 2).

2. We are illustrating our method with a three-way environmental distinction because this case is the simplest interesting case and it is easy to visualize. However, the method naturally generalizes to arbitrary numbers of classification categories.

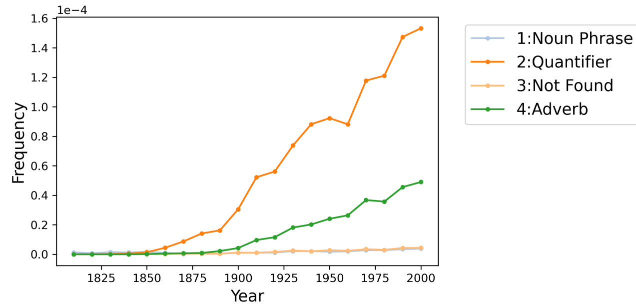

FIGURE 1 – Growth in overall frequency of *a lot*. Frequency scaled by number of words in corpus by decade. In the 20th century, the Noun Phrase and Not Found curves are overplotted.

Linear vs Higher Order Interpolation

We have tested this tabulation method on many words and phrases gleaned from the cleaned version of the COHA corpus (Alatrash et al., 2020)—495 million words covering American English over the period 1810-2009.[3] It turns out that, when examined using the method described above, the frequency profiles of many short, lexicalized expressions[4] have stayed more or less constant across the duration of the corpus. By contrast, the profiles of a few expressions (*a lot* among them) have changed dramatically. It is natural, then, to treat the average locations of the steady-state expressions as a backdrop against which we can view the trajectories of the changing expressions.

Figure 1 shows substantial growth in the frequency of *a lot* over time, especially in the innovating contexts, Quantifier and Adverb. In the current case, the Quantifier class captures proximity to the original use through ambiguity, as Quantifier constructions such as "a lot of clothes" are syntactically ambiguous between Noun Phrase and Quantifier. The exponential growth rate of the innovating contexts is approximately four times that of the progenitor context, Noun Phrase. By the last decade, the innovative instances outnumber the progenitors by a ratio of 52 :1. Such a large asymmetry is crucial to the coherence of the analysis method that we propose here : we are treating the probability vector as a profile of the evolving grammatical character of the innovative form. Our goal is to address the question of how and why the new form takes on its novel shape. If the relative frequencies of the two types remained the same over the course of the interval of study, then the trajectory would be a fixed point—the distribution would not change—so we would learn nothing about the innovation process by observing its character. Fortunately, many grammaticalization episodes exhibit a profile like that of *a lot* in this regard : there is massive growth of the innovative form

3. The online COHA corpus (Davies, 2008) seems to have been revised and extended since Alatrash et al. (2020) made a cleaned-up version of it in 2020, so the counts across the two versions do not coincide.

4. 1-2 words—we have not tested longer ones.

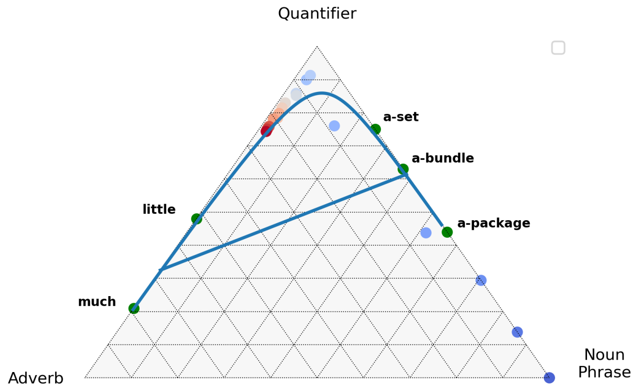

FIGURE 2 – Ternary diagram of *a lot*'s empirically estimated and hypothesized trajectories. Real data points are plotted inside the simplex and coded from earlier to later decades by colors from blue to red. Anchor points are plotted in green. Line segment : linear interpolation manifold. Curve : higher order interpolation manifold.

and zero or minimal growth of the progenitor.

Figure 2 shows the average locations of several expressions closely related in form and meaning to the initial state of *a lot* (*a package*, *a bundle*, *a set*) and also several related to the modern, adverbial meaning (*much*, *a little*). The vertex, *Noun Phrase*, means ⟨*a lot* unambiguously in one of its early nominal polysemies⟩. These include ⟨fate in life⟩, ⟨group of items packaged together for shipment or sale⟩, ⟨plot of land⟩, ⟨unruly crowd⟩, among others (4a). The vertex, *Quantifier*, means ⟨*a lot* with a following partitive *of*-phrase or with an implicit nominal reference set where the large quantity is plausibly emphasized so that a nominal or quantifier meaning is possible⟩ (4b). [5] The vertex, *Adverb*, means ⟨*a lot* conveying a high degree, where inclusion of a partitive *of*-phrase is either awkward, ungrammatical, or inconsistent with the intended meaning⟩ (4c). [6]

Figure 2 also depicts the *linear interpolation manifold* : the line segment between the NP-Quantifier and Quantifier-Adverb axes, which was obtained as the first principal component of the five anchor points. The baseline hypothesis (Linear Interpolation) claims that the state of *a lot* at any time between 1810 and 2009 is a weighted average of the NP-Quantifier anchors and the Quantifier-Adverb anchors. The change, on this view, is a progressive adjustment of the weighting, starting with all Noun Phrase, and ending with the particular mix of Adverb and Quantifier expressions that we observe. Including this hypothesis has several merits. First, the proximity thesis (a.k.a., the sneakiness characterization of De Smet (2012)) makes predictions about the order of change—predicting Noun Phrase → Quantifier → Adverb, and not Noun Phrase → Adverb → Quantifier. The linear interpolation hypothesis is unbiased with respect to order. Thus, if we detect a significant deviation from it all on one side, we can confirm one order over other possible orders. The linear interpolation hypothesis is closely related to an insightful model of statistical syntactic change, the Constant Rate Hypothesis of Kroch (1989) (see also Zimmermann, 2023). The Constant Rate Hypothesis claims that parametrically linked syntactic environments change in tandem—this amounts to linear interpolation in the space of logistic transforms of relative frequencies by meaning-environment. We believe this hypothesis has an important insight—but we note that it is not useful for explaining how novel syntactic forms get innovated because it only makes predictions about how structures change their frequency once they become grammatical. The nonlinearity of the Higher Order Interpolation hypothesis is the property that allows it to predict structural innovation as the consequence of preceding statistical change. The point of innovation is where the model diverges from an edge of the probability simplex. The linear movement along the edge of the probability simplex prior to innovation constitutes the statistical change that precedes innovation (QAQCG). To vindicate its claim, it is thus important to show that the relative probability vector can evolve nonlinearly.

The Higher Order Interpolation Model is defined by a 3-layer neural network (Figure 4). The inputs to the network are one-hot vectors uniquely encoding each backdrop term. For the purpose of modeling the 3-d backdrop distribution, a single hidden unit is sufficient because the class distributions for the progenitor (NP) and destination (Adverb) categories lie on locally 1-d manifolds that are relatively mildly independent. We can think of this as an Occam's Razor property of a hypothesized induction system for language : the learning system constructs a minimally elaborate model that can fit the data. The net input to each unit in the network is calculated by (Eq. 1) where a_i is a unit activation, b_i is the unit's bias, and w_{ij} is the weight from unit j to unit i. The hidden unit has hyperbolic tangent transfer function (Eq. 2). The target for each input is the probability profile of the input word.

$$net_i = b_i + \sum_i w_{ij} a_j \tag{1}$$

$$\tanh(net_i) = \frac{e^{net_i} - e^{-net_i}}{e^{net_i} + e^{-net_i}} \tag{2}$$

The output units have the softmax transfer function (Eq. 3).

$$o_i = \frac{e^{net_i}}{\sum_{j=1}^{n} e^{net_j}} \tag{3}$$

The model was trained by backpropagation until its error appeared to asymptote.

Statistical Evaluation Method

To compare the models, we need a statistical method of assessing fit to the data. Since the data lie on a probability

5. This classification has a lot of internal variety which we are not unpacking in the present paper. In addition to *a lot* grammaticalizing as an adverb, it is plausible that *a lot of* has grammaticalized into a plural quantifier over the interval we are considering (Selkirk, 1977).

6. For example, in (4c), it is awkward to include an *of*-phrase : ?? *Your cruiser's a lot of miles-per-hour slower than the Mooncat.*

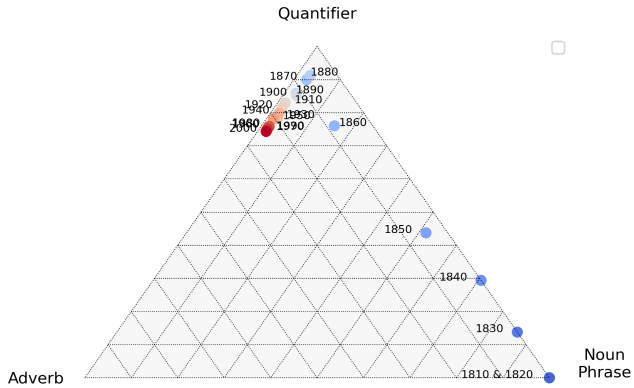

FIGURE 3 – Ternary diagram of *a lot*'s trajectory. Blue $\rightarrow$ Red shading encodes early $\rightarrow$ late decades, duplicating the information in the text year numbers which are sometimes hard to read because of overplotting.

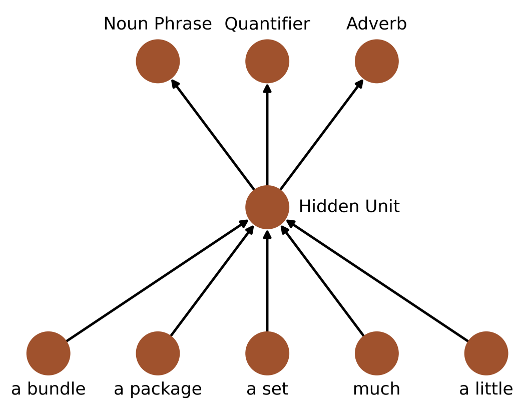

FIGURE 4 – Network architecture for *a lot* Higher Order Interpolation.

simplex, it is natural to consider noise with a Dirichlet distribution (Figure 5). Because each manifold is a trajectory over time, it can be specified by $\vec{o}(t) = [o_1(t), o_2(t), o_3(t)]$, where $o_i \in [0,1], \sum_i o_i(t) = 1$ and $t \in [0, \inf]$. N_t is the number of empirically observed tokens at a time point. The Dirichlet probability density function (pdf), *Dir*, is given by

$$Dir_t(x_1, x_2, x_3) = \frac{\prod_{i \in \{1,2,3\}} p_i^{(N_t o_i(t) - 1)}}{B(\vec{o}(t))} \quad (4)$$

where

$$B(N_t, \vec{o}(t)) = \frac{\prod_{i \in \{1,2,3\}} \Gamma(N_t o_i(t))}{\Gamma(\sum_{i \in \{1,2,3\}} N_t o_i(t))} \quad (5)$$

We optimized the Dirichlet pdf for each data point in the sample. Then, assuming a small constant area around each data point, dA, we approximated the probability of a data point as $dA \cdot Dir_{optimal}$ and computed the log probability of the sample assuming independent testing of the trajectory points.

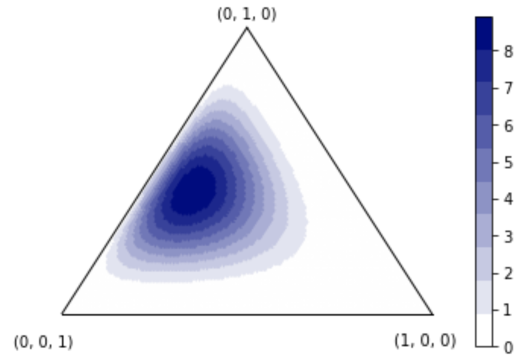

FIGURE 5 – A ternary diagram of the Dirichlet distribution where $\vec{o} = [0.2, 0.4, 0.4]$ and N = 10.

From this value, we computed Bayes factor, putting the Higher Order Interpolation Model in the numerator.

Case study : English *a lot*

We wrote code that allowed us to automatically classify almost every context in which *a lot* occurred in COHA. We did this by using Clean COHA tagging and defining preceding and following environments that matched regular expressions.

It turned out to be difficult to write regular expressions to handle the early period, when *lot* was mainly used as a noun. This is because interpretation of the pragmatic context is needed to identify the various polysemies of the early noun uses (discussed under Linear vs. Higher Order Interpolation above). Conveniently, these uses were already becoming quite outnumbered by the 1870s, and prior to 1870, there were only 264 instances, so we hand-coded these. This approach was crucial in capturing the Noun Phrase cases prior to 1870, which the automatic tagger would likely have misclassified as Quantifier uses. The resulting ternary plot is shown in Figure 3.

To assess the accuracy of our tagging system, we pulled a random sample of 200 instances of *a lot*. An informed research assistant hand-coded classifications for each of these, which we compared against the automatically tagged cases. We found that the tagger was 92% [7] accurate.

Data points in the years 1810 through 1849 were limited. When N is brought very low, the Dirichlet function morphs from a continuous unimodal distribution into a multimodal near-discrete distribution with the probability concentrated at one or more of the vertices of the simplex. Inferring such a stark classification on the basis of very tiny samples seems

7. Several cases were found to be difficult to classify for even the human tagger. In these cases, the human tagger was allowed to assign two tags and if either aligned with the automatic tag, it was counted as a match.

not meaningful so we dropped the first four decades from the sample (total of 71 data points dropped out of 44,200 or 0.16 percent of the sample).

Based on the remaining data, the Bayes Factor analysis yielded a value of $K = 1.91 \times 10^{27}$, indicating decisive support for the Higher Order Interpolation Model (Kass & Raftery, 1995).

Conclusions

We raised the hypothesis that languages taking part in historical grammaticalization episodes obey a Proximity Principle, and relatedly, that quantitative change anticipates qualitative (QAQCG). We proposed a formal method, Linear vs. Higher Order Interpolation for evaluating this set of claims. The method makes the assumption that a grammatical expression is usefully modeled as a point in a probability simplex, and that when its distribution changes over historical time, the point is following a continuous trajectory in the simplex. The corners of the simplex correspond to constructions which are motivated on the basis of linguistic analysis. The constructions associated with the corners need to be selected in such a way that all (or nearly all) instances of use of the expression under study that occur in a large historical sample are categorizable as belonging to one or another of the vertex types.

Several features of the outcome reported here are encouraging : (1) The method assumes that we can view the grammaticalization of a particular construction as occurring against the backdrop of relatively stable behavior of other constructions, providing a static reference frame with respect to which we can view the evolving form. This idea has been implicit in the conceptualizations offered by many historical linguists, but we are not aware of a previous formal method of measuring it or evaluating its validity. The present technique offers such a method and our case study of *a lot* is in line with the assumption that the anchor points all showed steady relative frequency profiles over time, even though some of them (*a package*, *a set*) grew substantially in per-word frequency. (2) Our method is designed to test the Proximity Principle (language "sneakiness") and QAQCG. In the case study, we obtained evidence that *a lot* conformed to the Proximity Principle's predicted ordering (Unambiguous → Ambiguous → Unambiguous).

Several other features give some pause : (1)' As we noted, the method will only be sensibly interpreted as modeling grammatical class migration if the innovative expression undergoes very rapid frequency growth relative to the progenitor, so that instances of the novel type substantially outnumber instances of the original type at the end of the measurement period. (2)' We had to work hard to achieve an accurate automatic classification scheme. We did this by writing just a few regular expressions at first, checking to see if the items culled by the method were properly classified and then diving into the (originally very large number of) instances that were "not found", and taking inspiration from these to devise additional regular expressions. It took many hours to complete this process and it turned out to be very important to work at it assiduously, because, we noted, a careless classification of the types could, in some cases, skew the results markedly. (3)' The anchor points were selected using intuition rather than a systematic procedure. We theorize that the entire language provides a stable backdrop for change, though the points closest to the changing form in semantic space have the most bearing. Due to the infeasibility of characterizing the entire stable backdrop, we used intuition to choose a set of stable constructions proximal to *a lot* syntactically and semantically. Future work should employ a more objective approach in defining the stable backdrop. (4)' We have only presented one case study. The method needs to be explored with other cases to see if it reliably works. More broadly, although property (1)' above implies that the data for the proposed kind of analysis are always going to be more scarce in the early years than the final years, it is desirable to find additional grammaticalization episodes where the years prior to the pronounced rise of innovative structures have more numerous absolute counts in the database.

To help with (2)', we put the concordance data that we used as the basis for the graphs and other analyses reported in this paper on an anonymous Open Science Framework (osf.io) site :

`https://osf.io/e2adp/?view_only=682345f4675a40bba7aea319db98ed1e`

Although these challenges are not trivial, our intuition is that they are, with careful effort, surmountable. If the current approach is vindicated empirically, then it points to an interpretation of what neural network models are saying about the nature of language encoding in contrast to classical symbolic models : they suggest that, despite appearances, syntactic structures may reside in a continuum. The continuum is a lower-dimensional manifold in a higher dimensional space defined by the constructional types that classical linguistic analysis has identified. Although the Higher Order Interpolation account does not specifically determine avenues of change (indeed, syntactic change is not constrained by structure alone, for dialects diverge), the analysis method presented here suggests that the account makes restrictive claims about the paths that change can follow, and these are statistically testable.

Références

Alatrash, R., Schlechtweg, D., Kuhn, J., & Im Walde, S. S. (2020). Ccoha : Clean corpus of historical american english. In *Proceedings of the twelfth language resources and evaluation conference* (pp. 6958–6966).

Andersen, H. (1987). From auxiliary to desinence. *Historical development of auxiliaries*, 21–52.

Bybee, J., Pagliuca, W., & Perkins, R. (1991). Back to the future. In E. Traugott & B. Heine (Eds.), *Approaches to grammaticalization, v. 2* (pp. 17–58). John Benjamins.

Bybee, J. L., Perkins, R. D. R. D., & Pagliuca, W. (1994). *The evolution of grammar : tense, aspect, and modality in the* *languages of the world*. Chicago : University of Chicago Press.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA : MIT Press.

Cinque, G. (1999). *Adverbs and functional heads : A cross-linguistic perspective*. Oxford University Press.

Craig, C. (1991). Ways to go in Rama : a case study in polygrammaticalization. In E. C. Traugott & B. Heine (Eds.), *Approaches to grammaticalization, v. 2* (pp. 455–492). John Benjamins.

Davies, M. (2008). *The corpus of contemporary american english (coca) : One billion words, 1990-2019.*

De Smet, H. (2012). The course of actualization. *Language (Baltimore)*, *88*(3), 601-633.

Devlin, J., Zbib, R., Huang, Z., Lamar, T., Schwartz, R., & Makhoul, J. (2014). Fast and robust neural network joint models for statistical machine translation. In *proceedings of the 52nd annual meeting of the association for computational linguistics (volume 1 : Long papers)* (pp. 1370–1380).

Diertani, C. E. A. (2011). *Morpheme boundaries and structural change : Affixes running amok* Thèse de doctorat non publiée. University of Pennsylvania.

Elman, J. L. (1990). Finding structure in time. *Cognitive science*, *14*(2), 179–211.

Hopper, P. J., & Traugott, E. C. (1993). *Grammaticalization, 2nd ed*. Cambridge, England : CUP.

Jespersen, O. (1917). Negation in english and other languages. *AF Høst*.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, *90*(430), 773–795.

Kroch, A. S. (1989). Reflexes of grammar in patterns of language change. *Journal of Language Variation and Change*, *1*(3), 199-244.

Lausberg, H. (1972). Romanische sprachwissenschaft : Iii. formenlehre. sammlung göschen 7199. *Berlin : De Gruffly*.

Lichtenberk, F. (1991). On the gradualness of grammaticalization. In E. C. Traugott & B. Heine (Eds.), *Approaches to grammaticalization, v. 1* (p. 37-80). John Benjamins.

Lord, C. (1976). Evidence for syntactic reanalysis : from verb to complementizer in Kwa. In *Papers from the parasession on diachronic syntax, chicago linguistic society* (pp. 179–91). University of Chicago Press.

Meillet, A. (1912). L'evolution des formes grammaticales. In *Linguistique historique et linguistic generale* (p. 13-48). Paris : Champion.

Norde, M. (2009). *Degrammaticalization*. Oxford University Press.

Roberts, I. (2010). Grammaticalization, the clausal hierarchy and semantic bleaching. *Gradience, gradualness and grammaticalization*, *45*, 73.

Roberts, I., & Roussou, A. (2003). *Syntactic change : a minimalist approach to grammaticalization*. Cambridge University Press.

Roumeliotis, K. I., & Tselikas, N. D. (2023, mai). Chatgpt and open-ai models : A preliminary review. *Future Internet*, *15*(6), 192. Consulté sur `https://www.mdpi.com/1999-5903/15/6/192` doi: 10.3390/fi15060192

Schmid, H.-J. (2020). *The dynamics of the linguistic system : Usage, conventionalization, and entrenchment* (First éd.). Oxford : Oxford University Press.

Selkirk, E. (1977). Some remarks on noun phrase structure. In A. Akmajian, P. Culicover, & T. Wasow (Eds.), *Studies in formal syntax* (p. 285-316). New York : Academic Press.

Sweetser, E. (1990). *From etymology to pragmatics*. Cambridge, England : Cambridge University Press.

Tabor, W. (1995). Lexical change as nonlinear interpolation. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the 17th annual cognitive science conference.* Lawrence Erlbaum Associates.

Traugott, E. C. (2017). Constructions in grammaticalization. *The handbook of historical linguistics*, 624–647.

Van Bree, C. (2004). Historische taalkunde. *Amersfoort : Acco*.

Zimmermann, R. (2023). An improved test of the constant rate hypothesis : late modern american english possessive have. *Corpus Linguistics and Linguistic Theory*, *19*(3), 323–352. Consulté le 2025-01-31, sur `https://doi.org/10.1515/cllt-2021-0038` doi: doi :10.1515/cllt-2021-0038